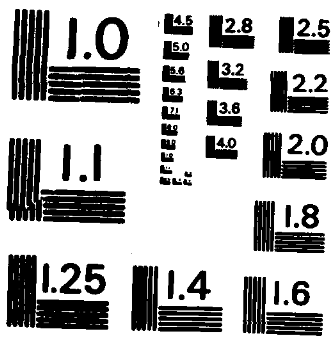
AD-A166 704 ADAPTIVE AIDING FOR HUMAN-COMPUTER CONTROL: EXPERIMENTAL STUDIES OF DYNAM. (U) ALPHATECH INC BURLINGTON MA N M MORRIS ET AL. JAN 86 TR-3 1/1

UNCLASSIFIED AAMRL-TR-86-005 F33516-82-C-0509 F/G 9/2 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

AAMRL-TR-86-005

AD-A166 704

ADAPTIVE AIDING FOR HUMAN-COMPUTER CONTROL: EXPERIMENTAL STUDIES OF DYNAMIC TASK ALLOCATION

NANCY M. MORRIS
WILLIAM B. ROUSE

Search Technology, Inc.

JANUARY 1986

Approved for public release; distribution is unlimited.

HARRY G. ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from Armstrong Aerospace Medical Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
Cameron Station
Alexandria, Virginia 22314

TECHNICAL REVIEW AND APPROVAL

AAMRL-TR-86-005

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

CHARLES BATES, JR.
Director, Human Engineering Division
Armstrong Aerospace Medical Research Laboratory

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AD-A166 704

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S): AAMRL-TR-86-005
6a. NAME OF PERFORMING ORGANIZATION: Search Technology, Inc. — 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: USAF, AFSC, AMD, Harry G. Armstrong Aerospace Medical Research Laboratory/
6c. ADDRESS (City, State and ZIP Code): 25-B Technology Park/Atlanta, Norcross GA 30092	7b. ADDRESS (City, State and ZIP Code): Wright-Patterson AFB OH 45433-6573
8a. NAME OF FUNDING/SPONSORING ORGANIZATION — 8b. OFFICE SYMBOL (If applicable): HEC	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F33615-82-C-0509
8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
62202F	7184	27	07

11. TITLE (Include Security Classification): See Reverse

12. PERSONAL AUTHOR(S): Morris, Nancy M. and Rouse, William B.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM ______ TO ______	1986 January	70

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.
05	05	
05	08	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Adaptive aids, adaptive systems, human-computer interface, computer aiding, adaptive expert systems, decision aiding, dynamic task allocation

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

(U) The results of two experiments investigating issues relevant to adaptive aiding are presented in this report. In the context of this effort, adaptive aids are those that allocate, partition, or transform tasks dynamically between a human and a computer in response to system or operator state in order to maximize system performance. In the first experiment subjects performed two tasks. In one condition an aid capable of performing one of the tasks was available. Subjects could activate the aid whenever they wished, but it only facilitated performance under certain conditions and, therefore, had to be used appropriately. As expected, subjects learned to use the aid and thereby improved performance relative to a no aid condition. In addition, the availability of the aid led to improved human performance even when the aid was not in use. From the results of this experiment, multiple regression models predicting subjects' performance under various task conditions were developed. These models were used as the basis for automating the task allocation decision in the second experiment. Subjects again performed two tasks, but
(continued on reverse)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☒ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Dr. John A. Forester	(513) 255-7576	AAMRL/HEC

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

ITEM 11. (Continued)

ADAPTIVE AIDING FOR HUMAN-COMPUTER CONTROL: EXPERIMENTAL STUDIES OF DYNAMIC TASK ALLOCATION (U)

ITEM 19. (Continued)

three aiding conditions were used: no aid available, aid under subject's control, and aid under computer control. The results showed that the aid was used more effectively when the decision was automated. However, the previously obtained benefit to unaided performance occurred only when the subject was in control of the task allocation decisions. The results of the study will ultimately be used as guidelines for the implementation of adaptive aiding in real systems.

SUMMARY

This is the third report in a series documenting an investigation of issues relevant to adaptive aiding. Within the context of this effort, adaptive aids are those that allocate, partition, or transform tasks dynamically in response to system or operator state in order to maximize system performance. The ultimate goal of this research is the identification of guidelines for the implementation of adaptive aids which can be useful to system designers.

In the experimental task environment developed for this research, subjects perform a subcritical compensatory tracking task while simultaneously identifying targets ("spotting") on a graphic display that moves down a CRT screen. A computer aid capable of identifying targets is sometimes available to perform the spotting task. The aid and spotting task are designed such that the relative superiority of human and computer may be expected to change over time; hence, the spotting task should be allocated dynamically to human or computer for best overall performance.

The results of two experiments in dynamic task allocation are presented in this report. In the first experiment, subjects performed both tasks with and without the spotting aid under various levels of tracking difficulty. Activation of the spotting aid was totally under subjects' control, and they were free to use the aid whenever they wished. Based on the results of this experiment, multiple regression models predicting subjects' performance in various task conditions were developed.

Adaptive Expert Systems.

These regression models served as the bases for automating the task allocation decision in the second experiment. Subjects again performed both tasks under three aiding conditions: no spotting aid available, spotting aid under subjects' control (manual aid), and spotting aid under control of the computer (automatic aid). Subjects' perceptions, opinions, and preferences regarding the tasks performed and aiding conditions were solicited via a questionnaire.

The results of these experiments may be summarized as follows:

1) Manipulations of task difficulty affected performance in anticipated directions; however, the interaction of spotting and tracking performance was rather weak.

2) Performance of the spotting task was affected by both current task difficulty and difficulty of the previous portion of the task.

3) Aiding improved overall spotting performance as expected.

4) The availability of the spotting aid led to improved human performance even when the aid was not in use.

5) Activation of the aid was more appropriate when the allocation decision was automated; however, the above benefit to unaided performance was realized only when subjects were in control of task allocation decisions.

6) Subjects occasionally overestimated the quality of their own performance.

7) Subjects wanted better performance from a human or computer assistant than they indicated would be acceptable from themselves.

PREFACE

This work was performed for the Human Engineering Division, Armstrong Aerospace Medical Research Laboratory at Wright-Patterson Air Force Base, in support of Project 2312-V2-33 (currently documented as 7184-27-07), Design Principles for Adaptive Decision Aids. The work was conducted by Search Technology, Inc. under subcontract to Alphatech, Inc., Contract Number F33615-82-C-0509.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

INTRODUCTION

This is the third annual report of a continuing effort devoted to investigation of issues relevant to adaptive aiding. As has been noted in other reports (e.g., Morris, Rouse, & Frey, 1985) adaptive aids are those that partition, allocate, or transform tasks dynamically in response to system or operator state, in order to maximize system performance. The concept of adaptive aiding is not new. However, it has recently gained popularity for two primary reasons. First, it is apparent that the complexity of existing and projected systems may easily exceed humans' abilities to deal with these systems. Second, advances in software and hardware technology have made implementation of the concept more feasible technically.

Although it is apparently feasible, implementation of adaptive aiding is not at all straightforward. There are a number of subtle and difficult issues which should be considered. For example, what should the aid's role in overall system operation be? How should the aid interact with the human? Is it possible for the aid to "understand" the human and supply assistance without overt communication from the human?

All too often in the past, decisions about the respective roles of humans and computers in engineering systems have been technology-driven. Tasks are automated because it is technically possible and economically feasible to automate them. The human is viewed merely as a component in the system, responsible for those odd jobs which are not yet automated, and functioning as a back-up system in case of failure of the automation.

In contrast, the guiding philosophy of the work reported here has been that the human is in charge of the system.* If humans are expected to assume responsibility for what happens to a system (particularly if something goes wrong), then they should be viewed as the <u>central</u> component in that system. From this perspective, automation should be used to enhance the human's role, not replace it. Thus, the overriding goal of this work is the provision of empirically-based guidelines for the use of automation to enhance human performance in engineering systems.

REVIEW OF PREVIOUS REPORTS

Work on this project conducted prior to this year consisted of three phases. First, issues relevant to adaptive aiding in general were outlined in the first-year report (Rouse & Rouse, 1983) and elaborated in the second-year report (Morris, Rouse, & Frey, 1985). As a result of this analysis, it became clear that investigation of all of the relevant issues would be a long and arduous process. Thus, in order to simplify matters temporarily, the immediate scope of the project was limited to investigation of adaptive allocation of tasks over time.

---

* This issue is elaborated in a working paper by Morris and Rouse (1985).

The second phase involved the development of a conceptual framework to serve as a means for organizing the large number of relationships viewed as relevant to dynamic task allocation and as a guide for selection of independent variables in experiments. This conceptual framework is described in detail in the second-year report, and continues to be useful. The third phase consisted of development of an experimental task environment designed to create conditions in which human and computer could interact. Pilot data were collected with the task environment to verify that characteristics of the environment affected performance in anticipated ways. The task environment and pilot research are also presented in the second-year report.

SCOPE OF THIS REPORT

A brief overview of the task environment is offered first as an aid to understanding the discussion of research which follows. The focus of this report is the presentation of two experiments in dynamic task allocation conducted within the context of the task environment. The first experiment was primarily an evaluation of the concept of adaptive task allocation, and task allocation was under subjects' control. In the second experiment, the effects of automating the task allocation decision were considered as well.

The results of these experiments offer interesting implications for implementation of adaptive aiding. These implications are discussed following the descriptions of the experiments. Considering these results in conjunction with the conceptual framework developed earlier, directions for future research efforts are also suggested.

DESCRIPTION OF TASK ENVIRONMENT

At present the task environment consists of two computer-based tasks which must be performed concurrently: a visual target recognition task, and a manual tracking task.

Target Recognition

Visual target recognition was chosen as one of the tasks in the scenario because of differences in the perceptual abilities of humans and computers. Humans readily impart meaning into what is seen, and are excellent at perceptual organization. Computers, on the other hand, have a great deal of difficulty analyzing scenes, but excel at figure rotation and template matching. Thus, humans should be better at identifying features in a meaningful scene, whereas computers should be better if the scene is a relatively homogeneous field of objects. The creation of conditions in which the human and computer should interact is accomplished by capitalizing upon these differences. The composition of the visual display changes over time, becoming more or less organizable.

When performing the target recognition task, subjects view a color graphic terrain display, which is illustrated in Figure 1. The terrain display depicts an intracoastal waterway with varying proportions of water. Water areas are colored blue. Also included in the display are green trees, tan ground, black buildings, white roads and parking lots, and cars and boats of assorted colors. To simulate flight over the terrain, the display pans down the CRT. Subjects are given the goal of identifying or spotting boats of a certain type which are in use in the waterway.

Targets may be identified only when they are in the spotting window defined by the heavy black horizontal lines. When the subject is identifying targets, identification is accomplished by using a mouse to position the cross-hair cursor on top of the target and then pressing a button on the mouse. When the button is pressed, a "+" appears on the screen and a tone is sounded by the terminal to acknowledge the action. Hits and false alarms are tallied in the upper left corner of the screen shown in Figure 1.

It is also possible for the computer "aid" to perform the spotting task. While the computer is identifying targets, the cross-hair cursor is not displayed. Actions on the part of the computer are acknowledged in the same manner as human actions, via symbols on the screen ("+") and tones from the terminal.

The relative performance of human and computer may be expected to vary over time due to the changes in the amount of water in the display. In light of the human's perceptual

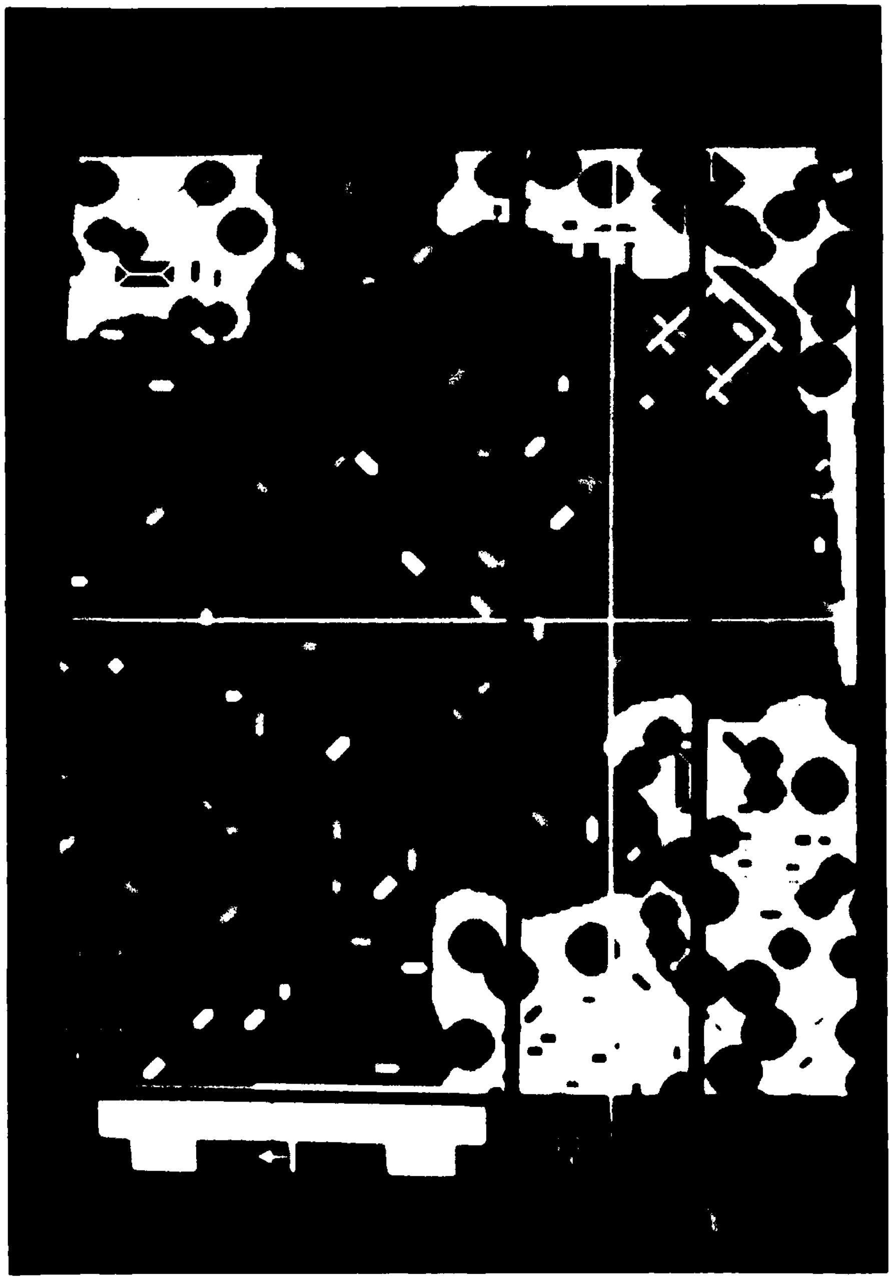


Figure 1. Task display.

abilities, this task should be easier for the human when the proportion of water in the spotting window is low (such as when flying over a narrow channel). This is because the human is able to organize the scene and automatically exclude a large portion (i.e., the land areas) from consideration.

The computer, on the other hand, is deficient in these organizational abilities, and scans the whole scene, identifying boats with a "template matching" (actually probabilistic) approach. As a result, the computer does not always differentiate land from water, and its false alarm rate increases with the proportion of land in the display. Thus, the human may be expected to excel when the proportion of water in the spotting window is low (i.e., over "channels"), and there is greater potential for the aid to excel when the proportion of water is high (i.e., over "bays").

Tracking

The second task employed is a subcritical compensatory tracking task, which is displayed in the upper left corner of Figure 1. The tracking display contains a green region flanked by yellow and red regions. The horizontal black line to the right of these regions moves up and down, and the arrow within the green region indicates the direction of the control input. The dynamic behavior of the tracking task is represented in equations 1 and 2.

$$z((n + 1)T) = r + c(z(nT)),\ T = 1/6 \text{ sec.} \qquad (1)$$

$$c = 1 + d/40 \qquad (2)$$

The tracking task is a modification of the tracking task developed by Jex, McDonnell, and Phatak (1966). Direction of movement of the controlled element is governed by the parameter "r", which toggles between $\pm$ maximum input. The value of the difficulty parameter, "d", is supplied by the experimenter at the beginning of an experimental run, and may have a value from 1 to 10.

The human's goal is to keep the black line within the green region by using bang-bang control via the space bar on the terminal keyboard. Should the moving pointer enter a red region, inputs from the mouse are disabled; hence, target identification is not possible unless the tracking task is also performed. When performing both tasks, the subject identifies targets with the right hand and tracks with the left.

Anticipated Need for Aid

With respect to the adaptive task allocation concept, it is possible to specify qualitatively when the computer should be used in this environment. First, the aid should be used if its potential target identification performance exceeds that of the human, an occurrence which is most likely when tracking is non-trivial and the terrain in the window is predominantly water. Second, the aid should be used to identify boats if the human's tracking performance degrades to an unacceptable level, which should also be related to both tracking difficulty and the amount of water in the window.

EXPERIMENT ONE

The primary goals of the first experiment were to demonstrate the utility of the adaptive task allocation concept and to gain insights into how people would make use of an aid capable of assuming control of some of their tasks. The degree to which use of the aid would reflect need for assistance (as indicated by performance decrement in unaided conditions) was of particular interest. It was also hoped that the performance data obtained would enable the development of a model of subjects' performance sufficient to allow automation of the task allocation decision.

METHOD

Independent Variables

Variables manipulated included terrain composition (and thus, spotting task difficulty), tracking difficulty, and availability of the aid. The panning speed of the target identification display was held constant, so that the time required for an object to traverse the spotting window was approximately 10 seconds. Four levels of the tracking task difficulty parameter were used: 1, 3, 5, or 7. If the "time constant" for the tracking task is defined as the number of seconds for the cursor to travel from the center of the display to one of the edges of the green area given no control input, the time constants for the above difficulty levels were 2.699, 0.920, 0.564, and 0.410 seconds, respectively.

As discussed earlier, spotting task difficulty was a function of the amount of water in the spotting window, and thus varied as the terrain display panned down the screen. When the spotting aid was available, it was activated and deactivated manually by subjects as they desired. To activate the aid and turn control of the spotting task over to the computer, subjects positioned the cross-hair cursor used in the spotting task on top of the word "AID" (displayed on the left side of the screen), and pressed a button on the mouse. Aid deactivation was also accomplished by pressing a button on the mouse.

Subjects and Experimental Procedure

Ten volunteers from the AMRL subject pool served in the experiment, and were paid for participating. Five had no prior experience with the task environment and served in 10 sessions each (5 without the aid, followed by 5 with the aid available). The remaining 5 subjects had already received some practice on the tasks (including use of the aid), and served in 3 unaided and 4 aided sessions. The treatment received by the latter group of subjects was the same as in the last sessions of the former group (i.e., sessions 1-3 and 4-7 for the latter group were the same as sessions 3-5 and 7-10, respectively, for the former group).

Each session consisted of one period of approximately 5 minutes of spotting (i.e., one pass over the terrain display) under tracking difficulty of 1, followed by two periods of spotting under each of the other levels of tracking difficulty (for a total of seven 5-minute periods). Order of tracking difficulty for the last six periods was pseudo-random. Since these periods were self-started via a carriage return at the terminal keyboard, subjects were able to rest between periods whenever they wished.

RESULTS

Data from the last three unaided and last three aided sessions were analyzed via a variety of statistical procedures. Only those effects which were statistically significant are reported. When determining which effects were significant, the criterion for significance was a p value of .05 or less; most values of p for the results reported were considerably lower.

Several dependent measures were examined, some of which are presented here. The primary performance measure for the tracking task was rms tracking error. For the spotting task, the primary measure of performance was hits, defined as percent of targets present which were identified. Latency of hits, or time elapsed from the entry of a target into the spotting window until it was identified, was another measure of spotting performance. The latency measure is discussed in conjunction with some of the multiple regression analyses.

False alarms on the spotting task were also examined, and some significant differences were noted (e.g., as expected, the aid made more false alarms than did humans). False alarms are not discussed, however, because the frequencies of false alarms under all conditions were very low. Percent of each terrain segment which was exposed to the aid was the primary measure of aid use.

Differences Between Task Conditions

First, differences in performance associated with task conditions were assessed via analysis of variance with repeated measures. Initially, four factors were included in the analyses: aiding (no aid vs. aid), tracking difficulty (4 levels), percent water in the terrain display (6 levels), and session (3 levels). The results of these analyses were rather confusing, however, and detailed examination of the data suggested that alternative factors would be more appropriate. Analyses including the following factors produced more satisfactory results: aiding (no aid vs. aid), tracking difficulty (4 levels), current terrain composition (low vs. high percent water currently in the spotting window, or "channel" vs. "bay"), and previous terrain composition (low vs. high percent water in the terrain segment which just exited the spotting window).

In the following presentation of the effects of task variables on human performance, only unaided sessions were considered. Several significant effects were noted.

Effects of task parameters on performance. There was a strong effect of tracking difficulty on rms error on the tracking task (ranging from 26.41 with the easiest level of tracking to 41.99 with the most difficult). There were also strong effects of terrain type on spotting performance (i.e., hits). Interestingly, there was an interaction of current and previous terrain type on spotting performance, shown in Figure 2. When spotting over channels, previous terrain had little effect on hits (89.09% when the previous terrain segment also contained a channel, compared to 86.06% when the previous terrain included a bay). However, when spotting over a bay, the effects of the previous terrain type were quite noticeable (65.62% hits when the previous terrain contained a channel, vs. 45.94% if there had been a bay in the previous terrain).

Tradeoffs in performance of two tasks. Prior to conducting the experiment, it was expected that performance of the tracking and spotting task would interact, with good performance on one achieved at the expense of performance on the other. However, these effects proved to be rather weak. There was a small but significant increase in rms error accompanying increases in the amount of water in the display (ranging from 34.71 to 36.02). There was also a small decrement in hits on the spotting task

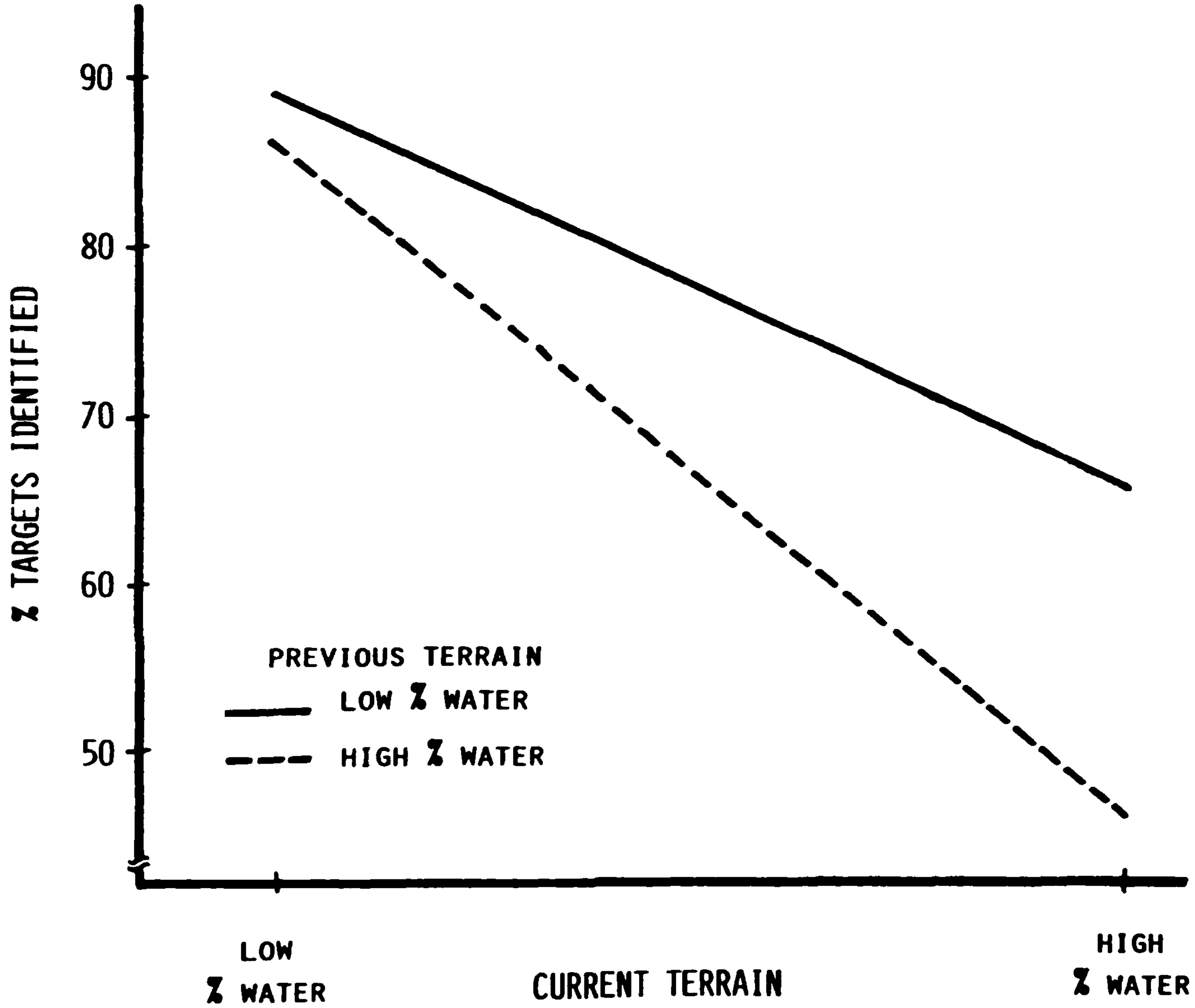


Figure 2. Unaided spotting performance as a function of terrain composition.

associated with increasing tracking difficulty (from 74.06% to 68.96%).

Effects of aiding. Sessions in which the aid was available were compared to unaided sessions in order to assess the effects of having an aid. When the aid was available to perform the spotting task, rms error on the tracking task was lower (31.79 vs. 35.54 without the aid). This difference was greater when spotting over a bay (a difference of 5.05, compared to 2.45 over a channel), probably because subjects tended to use the aid when over open water. (How subjects used the aid is discussed later.) These results are presented graphically in Figure 3.

As may be seen in Figure 4, there was also an improvement in spotting performance when the aid was available (89.28% hits vs. 71.68% without the aid). Compared to unaided sessions, there was less decrement in spotting performance when the percent of water in the current terrain was high (82.97% and 81.58% for low and high percent water in previous terrain types, respectively, compared to 65.62% and 45.94% noted earlier). Additionally, in contrast to the unaided sessions, there was no decrement in spotting performance accompanying increases in tracking difficulty.

Subjects' spotting performance over terrain segments when the aid was available but turned off (i.e., less than 50% exposed to the aid) was compared to their performance over identical terrain segments when the aid was not available. This comparision revealed that subjects identified approximately 10% more targets

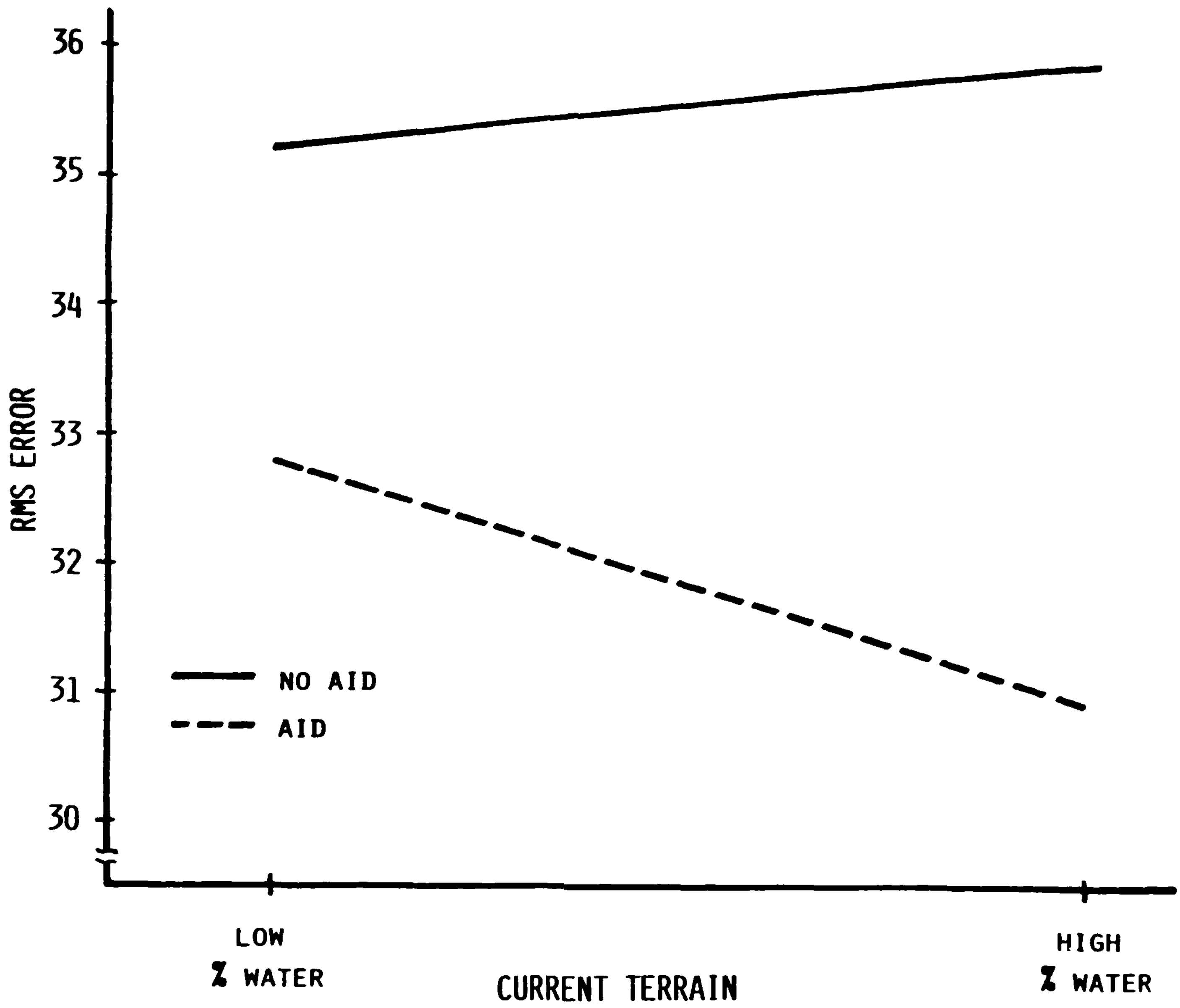


Figure 3. RMS tracking error when spotting over bays vs. channels, with and without spotting aid available.

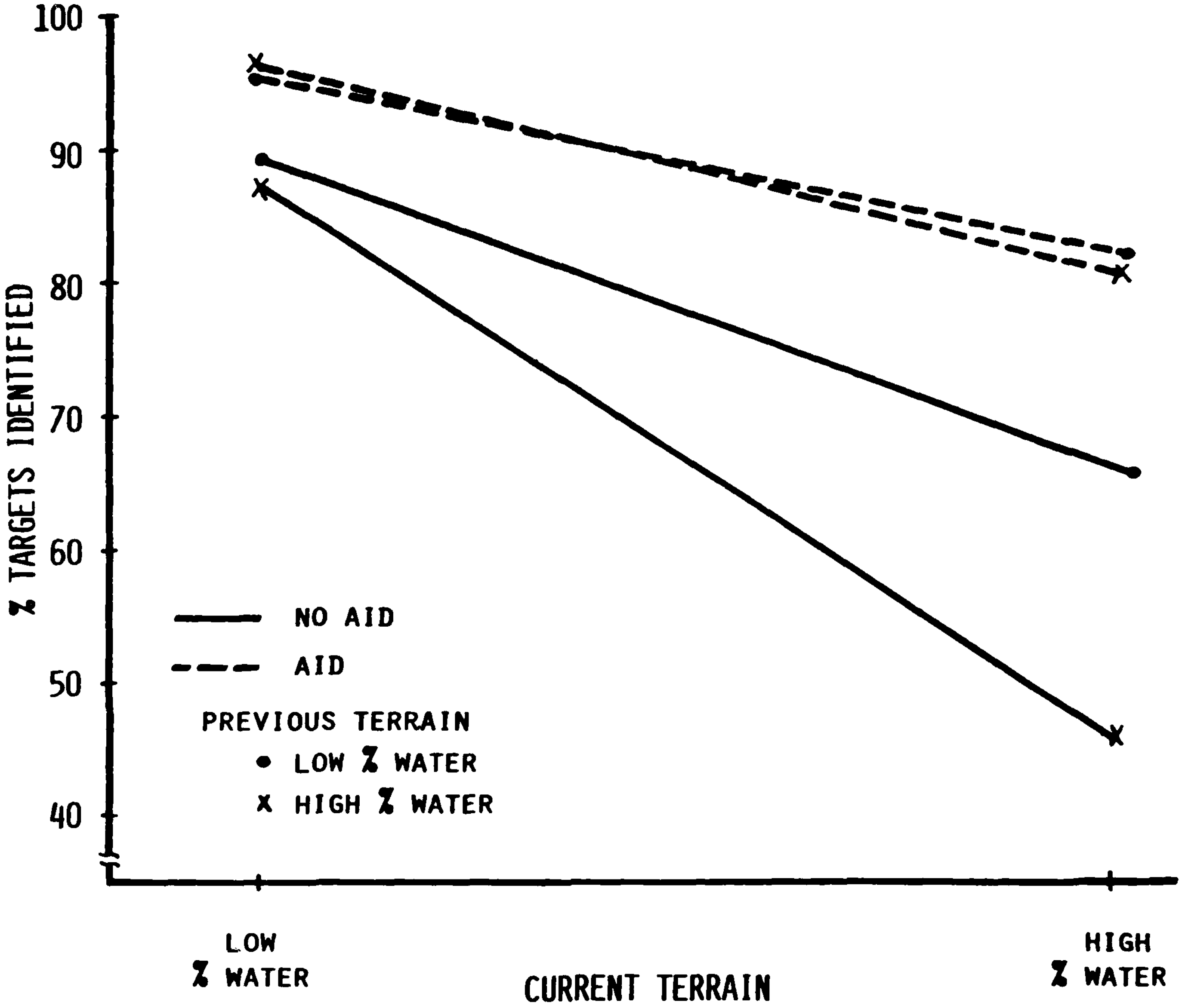


Figure 4. Spotting performance as a function of terrain composition, with and without spotting aid available.

in aided sessions than in comparable conditions in the unaided sessions. Hence, it would seem that aiding subjects during the more difficult portions of their task enabled them to perform better on the unaided portions. This result was interpreted as merely suggestive, however, because of the design of the experiment; aided sessions always occurred after unaided sessions, so better performance could be the result of learning rather than aiding.

An analysis of performance in each of the three unaided and three aided sessions revealed significant effects of both session and aid, with sessions 2 and 3 in each condition better than session 1 but no different from each other. That at least some of this improvement in performance could be a side effect of aiding seemed plausible for two reasons. First, the fact that performance did not improve between sessions 2 and 3 suggested that performance had stabilized somewhat. Second, the average difference between aided and unaided sessions was greater than the average difference within aiding conditions (10% vs. 3%). Nevertheless, the possibility that this effect was due to learning could not be ruled out. This issue was addressed further in the next experiment.

Prediction of Performance

To enable a finer-grained analysis of subjects' performance, multiple regression equations were determined for each subject individually and for all subjects combined. It was also hoped that multiple regression equations could be used as online models to serve as a basis for automated adaptive task allocation in the future. Selection of predictor variables was based upon the results of analysis of variance. Predictor variables used were tracking difficulty (the time constant of the controlled dynamics), current terrain type (percent water), previous terrain type, and previous x current terrain type.

Effects of task parameters and performance tradeoffs. Not surprisingly, the results of multiple regression were consistent with the results of analysis of variance in that manipulations of the difficulty of one of the tasks affected performance of that task but had little effect upon performance of the other task. When predicting rms error, coefficients for tracking difficulty were significant and large; coefficients for terrain type were large when hits were predicted. In the few cases where coefficients for the "opposite" task were significant (e.g., a significant coefficient for tracking difficulty when predicting hits), the magnitude of those coefficients was small.

The performance measures of rms error, hits, and latency of hits were included in the set of predictor variables and additional regression equations were determined. Coefficients for performance measures were significant for only a few subjects, and the maximum contribution of performance measures to variance explained was only three percent.

Differences in goodness of fit of models. Three different approaches to regression were examined. First, regression equations were fit to mean performance (i.e., collapsed across multiple occurrences of each combination of tracking difficulty and past x current terrain composition). Generally, the fits of these equations were rather good. When predicting mean rms error, inclusion of all subjects' means produced an R of .79; for individual solutions, R ranged from .73 to .91. Prediction of mean percent hits was slightly better: R for the group was .86, and individual R's ranged from .84 to .94.

When regression equations were fit to raw data, it was not surprising that the fits were not nearly as good, with R's ranging from .48 to .64 for rms error, and from .57 to .82 for hits. When predictions based on the equations derived from means were compared to the raw data, fits were approximately the same as for the regression on raw data (for rms error, R ranged from .39 to .64; for hits, R once again ranged from .57 to .82).

Use of aid. In order to examine how subjects used the aid, the same set of predictor variables was used to predict the percent of each terrain type which was exposed to the aid (i.e., tracking difficulty, current terrain type, previous terrain type, and previous x current terrain type). When all subjects were included in the regression based on means, R was found to be .76; individual R's ranged from .58 to .97. The aforementioned regressions on raw data and predictions of raw data based on means were also performed, with results similar to those mentioned earlier. Individual R's resulting from regressions on raw data ranged from .28 to .90; when predicting raw data from means, R's ranged from .29 to .90. When examining the results of the individual regression equations, it was noted that the fits for most subjects were very good, with only two or three subjects having low R's.

Predicted Need vs. Actual Use of Aid

The primary purpose of predicting performance via multiple regression was to enable online decision making in the next experiment. The expected quality of these online task allocation decisions was evaluated by 1) determining which partner would have been in control of the spotting task under each task condition if the allocation decision had been automated, and 2) examining actual performance to determine if the task allocation decision would have been appropriate. Specifically, individual regression equations derived from means were used to predict rms error and hits for each subject in each of the 60 task conditions (4 levels of tracking difficulty x 15 combinations of previous and current terrain type which appeared in the problems used in this experiment). These predictions were then compared to expected performance on the part of the aid, and judgments as to when the aid should be used by each subject were made, based on the anticipated superiority of human or computer in each condition. Similar judgments based on comparison of the aid's expected performance to actual mean performance achieved by subjects in each condition were also made, and discrepancies in the two sets of judgments were noted. Discrepancies between predicted need for the aid and actual use of the aid were also examined.

The original intention was to base decisions regarding need for the aid on performance of both the spotting task and tracking task, with the idea that the aid was needed if performance on either task degraded. A criterion of 22 or less was selected as acceptable rms error (a score which would be achieved if the direction of movement of the tracking indicator was always reversed at the green-yellow border on the tracking display) and each subject's spotting performance was judged relative to the predicted performance of the aid. (The aid's performance was determined by subtracting expected false alarms from expected hits.) When decisions were made on this basis, however, the result was that the aid was needed almost all of the time, because rms error was <u>very</u> rarely below the criterion level. Therefore, only spotting performance was considered in the following discussion.

When subjects' use of the aid was compared to predicted need for the aid, it was found that seven of the ten subjects' average usage agreed with predicted need more than 90% of the time, whereas three of the subjects used the aid in accord with predicted need only 60-72% of the time. A detailed analysis of discrepancies revealed that, for almost all subjects, most discrepancies resulted from subjects' performing the spotting task themselves rather than letting the aid do it as it was predicted they should.

An examination of performance in the discrepant conditions revealed differences between the two groups of subjects. For the seven who agreed with predictions very closely, it was noted that most of the time their performance was superior to the aid's and that they were right to do the spotting themselves. In contrast, when the other three subjects did not use the aid according to predicted need, their performance was worse than the aid's most of the time and the aid should have performed the spotting task. One of these three subjects also had the aid perform the spotting task when it was predicted that he should perform the task instead; in his case, the prediction was right in every such discrepancy.

Summary of Results

Interpretation and discussion of the results of this experiment are postponed until the results of Experiment Two are presented. The following list is a brief summary of the results described thus far.

- Regardless of tracking difficulty, tracking error was usually greater than the acceptable level indicated to subjects at the beginning of the experiment. Tracking error (rms) increased as the controlled element was more unstable.

- Spotting performance (hits) when no aid was available was worse over bays than over channels. There was also a "carry-over" effect of the previous terrain. For example, spotting over a channel was worse if the previous terrain was part of a bay rather than a channel.

- Contrary to expectations, strong tradeoffs in performance of the two tasks were not very evident. Tracking error was slightly higher when spotting over a bay rather than a channel, and spotting performance was slightly worse as tracking difficulty increased.

- When the aid was available to do the spotting task, rms tracking error was lower, particularly when spotting over water (i.e., when the aid was typically in use).

- When the aid was available, more hits were achieved and the detrimental effects of terrain composition upon spotting performance were reduced. There was also no effect of tracking difficulty upon percent hits.

- When the aid was available, subjects identified more targets when the aid was turned off than in comparable terrain segments from passes in which no aid was available. Because of the experimental design, the possibility that this effect was due to learning could not be ruled out.

- Regressions of task characteristics on performance measures were consistent with the results of analysis of variance. Inclusion of behavioral measures in regression solutions did not substantially increase the predictive ability of those solutions.

- For most of the subjects, average use of the aid corresponded to predicted need quite closely. In most discrepant cases, subjects used the aid less than predictions indicated they should have; usually, subjects were correct in these discrepancies. Only one subject used the aid more than suggested by predictions, and his use of the aid was usually inappropriate.

EXPERIMENT TWO

The primary goals of the second experiment were to investigate the effects of automating the task allocation decision on performance and to gain insights into subjects' opinions and preferences with respect to automated decision making. The effects upon subsequent human performance of having an aid perform portions of one's task were also of interest, in light of the results of the first experiment. Finally, information relative to subjects' abilities to estimate the quality of their own performance was sought, since the accuracy of humans' perceptions is viewed as a primary factor influencing the quality of task allocation decisions on the part of the human. This factor occupies a central role in the conceptual framework noted earlier.

METHOD

Modification of Experimental Task Environment

Some modifications to the experimental task environment were made to allow for automation of task allocation decisions. The bases for these decisions were the regression models developed from data obtained in Experiment One. Referring to the discussion of the results of the first experiment, regressions based on means were used; recall that the parameters of these models included tracking difficulty and terrain type. Individual equations for each subject serving in Experiment One were available, as well as one group equation based on all ten of the subjects in the first experiment.

Since both automated and human decision making were to be included in the experiment, the display was altered to indicate which mode of decision making was currently in effect. More specifically, the word "AID" to the left of the terrain display was changed to "MANUAL AID" when the human was in charge of task allocation, and "AUTO AID" when task allocation decisions were automated. Functioning of the manual aid was identical to that described in the previous experiment. When automatic aiding was in effect, the following warning was given 5 seconds before control of the spotting task was to be transferred to the computer or vice versa: the words "AUTO AID" blinked a few times (via reverse video) and a warning tone was sounded by the terminal.

Questionnaire

A questionnaire to obtain subjects' opinions, preferences, etc., was developed. (This questionnaire may be found in the Appendix.) Questions asked subjects to judge the quality of their own performance and specify their criteria for "acceptable" performance in themselves and in human and computer assistants. Their opinions about the approaches to aiding used in the experiment were also sought, as well as preferences about assistance in general.

Independent Variables

The independent variables in this experiment were spotting task difficulty and aid availability. Spotting task difficulty was manipulated via terrain composition as in the first experiment; in fact, the same terrain displays were used. Three levels of aid availability were used: no aid available, manual aiding, and automatic aiding. As in Experiment One, the panning speed of the spotting task was held at a constant speed of approximately 10 seconds for a target to traverse the spotting window. In light of the failure of tracking difficulty to produce any substantial differences in spotting performance in the first experiment, tracking difficulty was held constant at 3 (i.e., a time constant of 0.92 seconds).

Subjects and Experimental Procedure

Ten persons from the AMRL subject pool served as paid volunteer subjects. Of the ten, eight had served in the first experiment, and a ninth had prior exposure to the task and manual aiding. The tenth person, who had no previous experience with the task environment, served in two practice sessions designed to provide comparable exposure to the tasks. These practice sessions consisted of one unaided session followed by one session with the manual aid available; tracking difficulty was varied in the practice sessions as in Experiment One. With the exception of these two practice sessions, all subjects received identical treatment as described below.

Subjects served in two sessions each. Each session began with one pass over the terrain display with no aid available and tracking difficulty of 1. This was followed by two passes under each of the three levels of aid availability (i.e., no aid, manual aid, and automatic aid), for a total of seven passes over the terrain per session. The two passes under a given aiding condition were presented as a block and the order of presentation of blocks was pseudo-random, counterbalancing order of presentation between subjects and within subjects across sessions as much as possible. As in the first experiment, each pass was self-started by entering a carriage return at the terminal keyboard. Each subject filled out the questionnaire at the end of the last session.

For each of the eight subjects who participated in the first experiment, automatic task allocation decisions were based on the individual regression model derived for that subject. Automatic task allocation decisions for the two subjects for whom no individual models were available were based on the group regression model.

RESULTS

Performance and questionnaire data were analyzed via a variety of statistical techniques. As with the first experiment, the criterion for statistical significance was a $\underline{p}$ value of .05 or less. In the following discussion, all reported differences were statistically significant. Dependent measures of task performance were those discussed in the first experiment: rms tracking error, hits (percent targets identified), false alarms, and latency of hits. Percent of each terrain segment exposed to the aid was once again the primary index of aid usage.

Differences Between Task Conditions

Performance data from the second session only were analyzed via analysis of variance with repeated measures. Factors in the analyses included aiding (no aid, manual aid, auto aid), previous terrain composition (low percent water vs. high percent water), current terrain composition (low percent water vs. high percent water), and previous x current terrain. The following effects were noted.

Effects of task parameters on human performance. Human performance of the tracking and spotting tasks was consistent with that observed in the first experiment. There was an interaction of previous and current terrain composition in their effects upon spotting performance, as indicated by percent targets identified (hits). When the amount of water in both the previous and current terrain was low, subjects identified 88.78% of the targets; however, if the previous terrain segment contained a high proportion of water, 84.98% of the targets were spotted. When the amount of water in the current terrain segment was high, subjects achieved 68.36% hits if the previous terrain contained little water; finally, if both previous and current terrain contained high percentages of water, only 46.12% of the targets were identified. These results are quite close to those obtained in the first experiment, as depicted in Figure 2.

Minimal effects of spotting task difficulty on tracking performance were observed, which was also consistent with the results of Experiment One. When the amount of water in the previous terrain segment was low, rms tracking error was slightly lower (31.73 vs. 33.33 when the amount of water in the previous terrain was high). The effects of current terrain composition on rms error were similar (33.19 over narrow channels vs. 35.08 over open water).

Effects of aiding on system performance. Overall system performance improved when an aid was available. The following presentation of significant effects may be better understood if one recalls that the aid, when available, generally performed the spotting task over open water, and subjects spotted target boats in channels. This issue will be elaborated later, in discussing the results.

Tracking error was greater when no aid was available (34.14) than with manual aid (32.35) or auto aid (31.10); there was no differential effect of type of aid on rms error. Generally, the effect of aiding was to reduce (in fact, reverse) the impact of spotting difficulty on tracking performance. In aided sessions, rms error was lower if the amount of water in the previous terrain was high (30.64 vs. 32.80 for little water in the previous terrain), and if the amount of water in the current terrain was high (30.39 vs. 33.06 over channels).

More targets were identified when an aid was available than in unaided passes (87.72% and 89.95% with manual and automatic aids, respectively, vs. 72.06% with no aid). There was also an interaction with type of aid and composition of the current terrain segment, shown in Figure 5. When the manual aid was available, a larger percentage of targets was identified over channels than over open water (91.50% vs. 83.94%); however, when automatic aiding was in effect, the percentage of targets identified over channels was less than over open water (88.49% vs. 91.42%).

More false alarms occurred when an aid was available (0.55 and 0.58 per terrain segment with manual and automatic aids, respectively, vs. 0.20 with no aid available). When subjects were in control of aid activation (i.e., manual aiding), more false alarms occurred over channels than with either no aid or automatic aiding (0.32 vs. 0.12 and 0.14 for no aid and automatic aiding, respectively). However, when spotting over open water, there was no significant difference in false alarms between manual and automatic aids (0.78 with manual, vs. 1.02 with automatic), both of which were greater than false alarms with no aid available (0.28). These results are presented graphically in Figure 6.

Activity of manual vs. automatic aids. Some differences between manual and automatic aiding have been noted in the above discussion of hits and false alarms across aiding conditions. Insights into the reasons for these differences, as well as how subjects made use of the manual aid, may be gained from examining

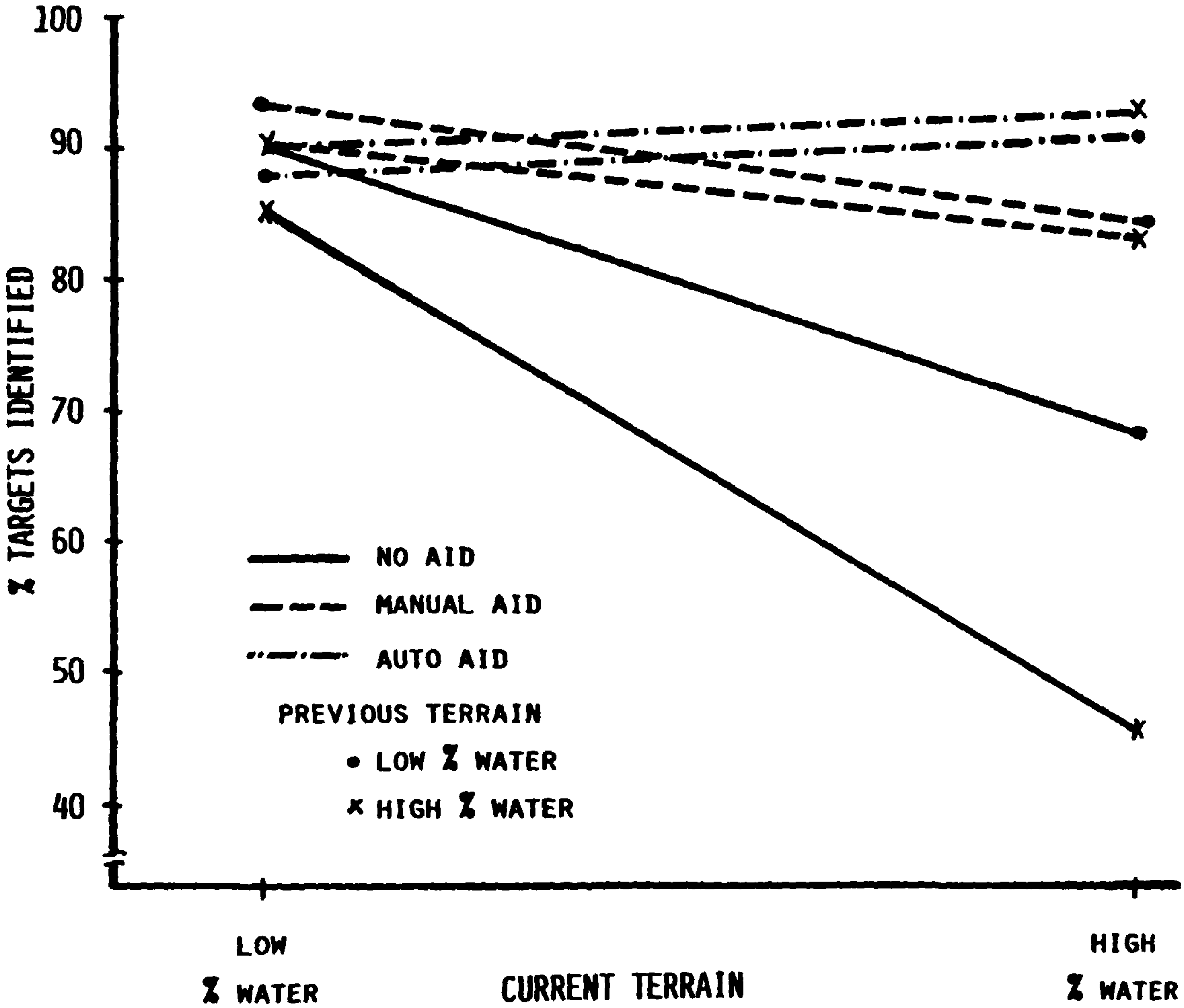


Figure 5. Spotting performance as affected by terrain composition under three aiding conditions.

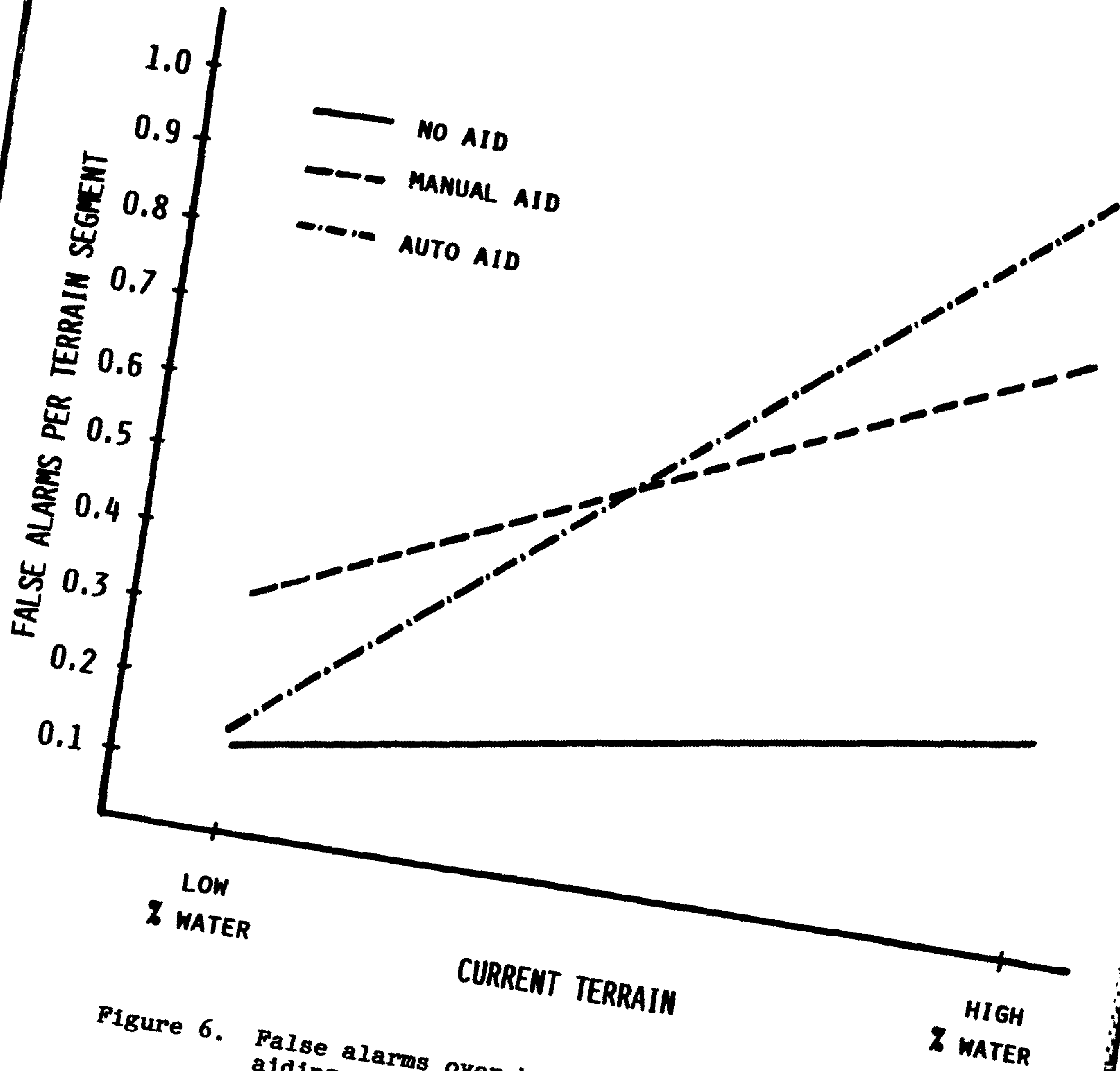


Figure 6. False alarms over bays and channels under three aiding conditions.

patterns of aid usage. Only the two aided conditions were included in analysis of the following effects.

There were differences between manual and automatic aiding in the percent of each terrain segment which was exposed to the aid, as shown in Figure 7. When automatic aiding was in effect, terrain segments containing little water were exposed to the aid less (0.52% vs. 13.17% with manual aiding), and terrain segments containing a great deal of water were exposed to the aid more (86.28% vs. 66.54% with manual aiding). Furthermore, the automatic aid initiated transfer of the spotting task less often over channels (0.0 interactions with the subject per terrain segment, vs. 0.07 with manual aid), and more frequently over open water (0.68 vs. 0.49 with manual aid). Examination of false alarms made by the aid (presented in Figure 8) revealed that the automatic aid made fewer false alarms over channels (0.0 vs. 0.20 for the manual aid), and more false alarms over open water (0.91 vs. 0.63 for the manual aid).

There was an interaction of previous and current terrain composition on percent hits made by the two types of aid, which is presented in Figure 9. When the previous terrain segment contained little water, the automatic aid made fewer hits (0.0% vs. 1.17% by the manual aid when the current terrain contained little water, and 0.0% vs. 13.83% by the manual aid when the current terrain contained a large amount of water). Hits achieved by the two aids were approximately the same in the high previous, low current condition; however, when the amount of water in the

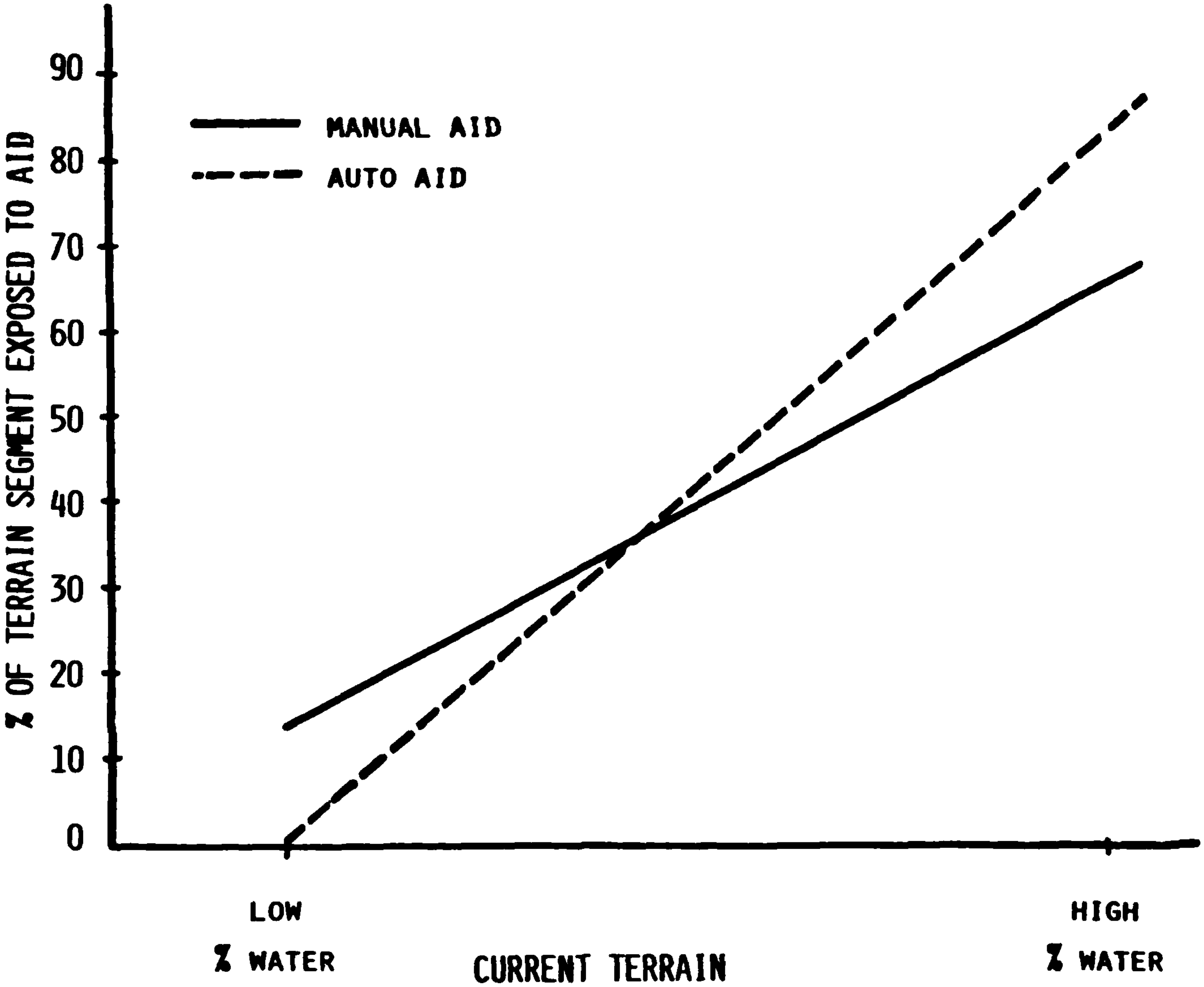


Figure 7. Percent of terrain exposed to each type of aid.

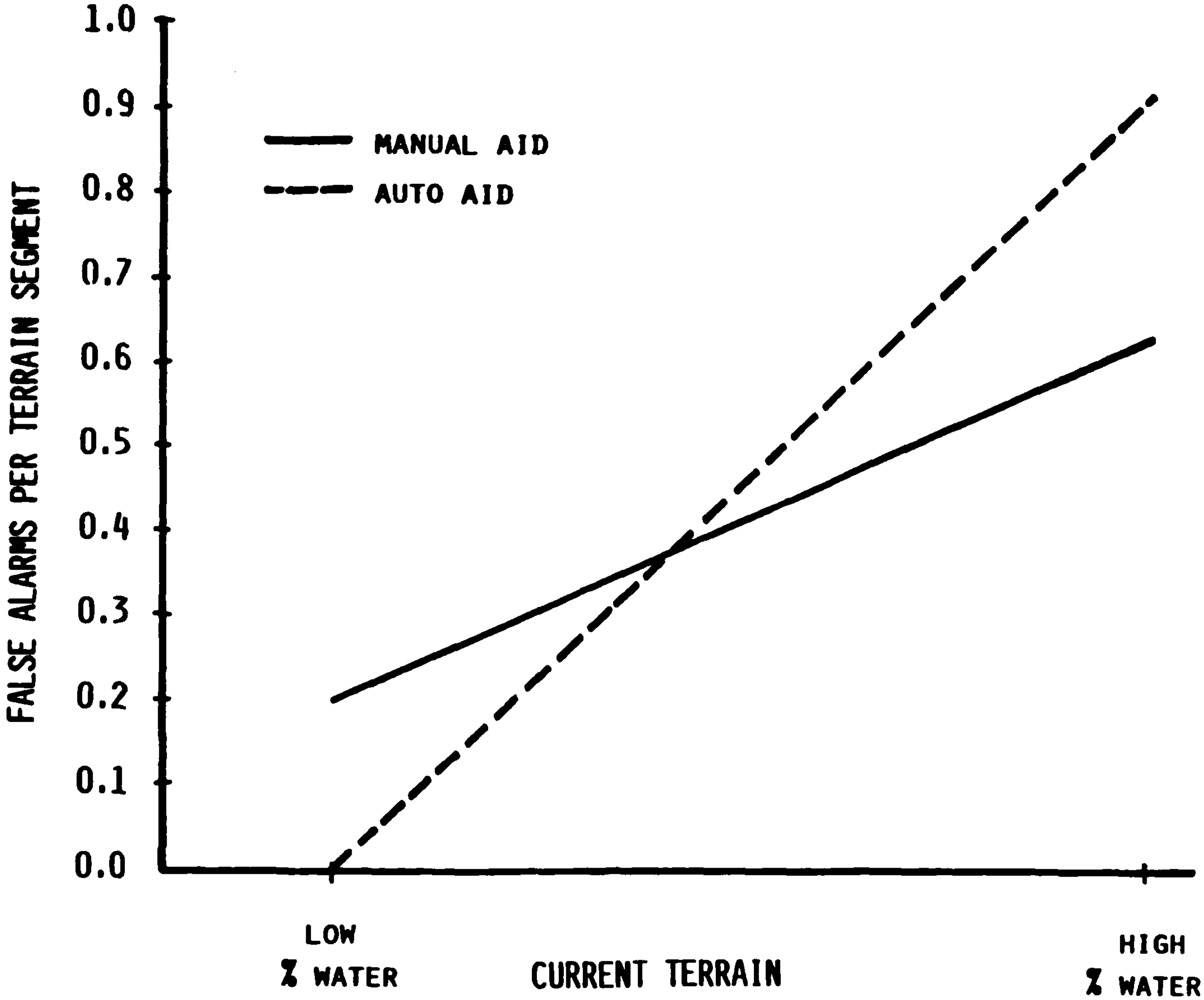


Figure 8. False alarms made by each type of aid.

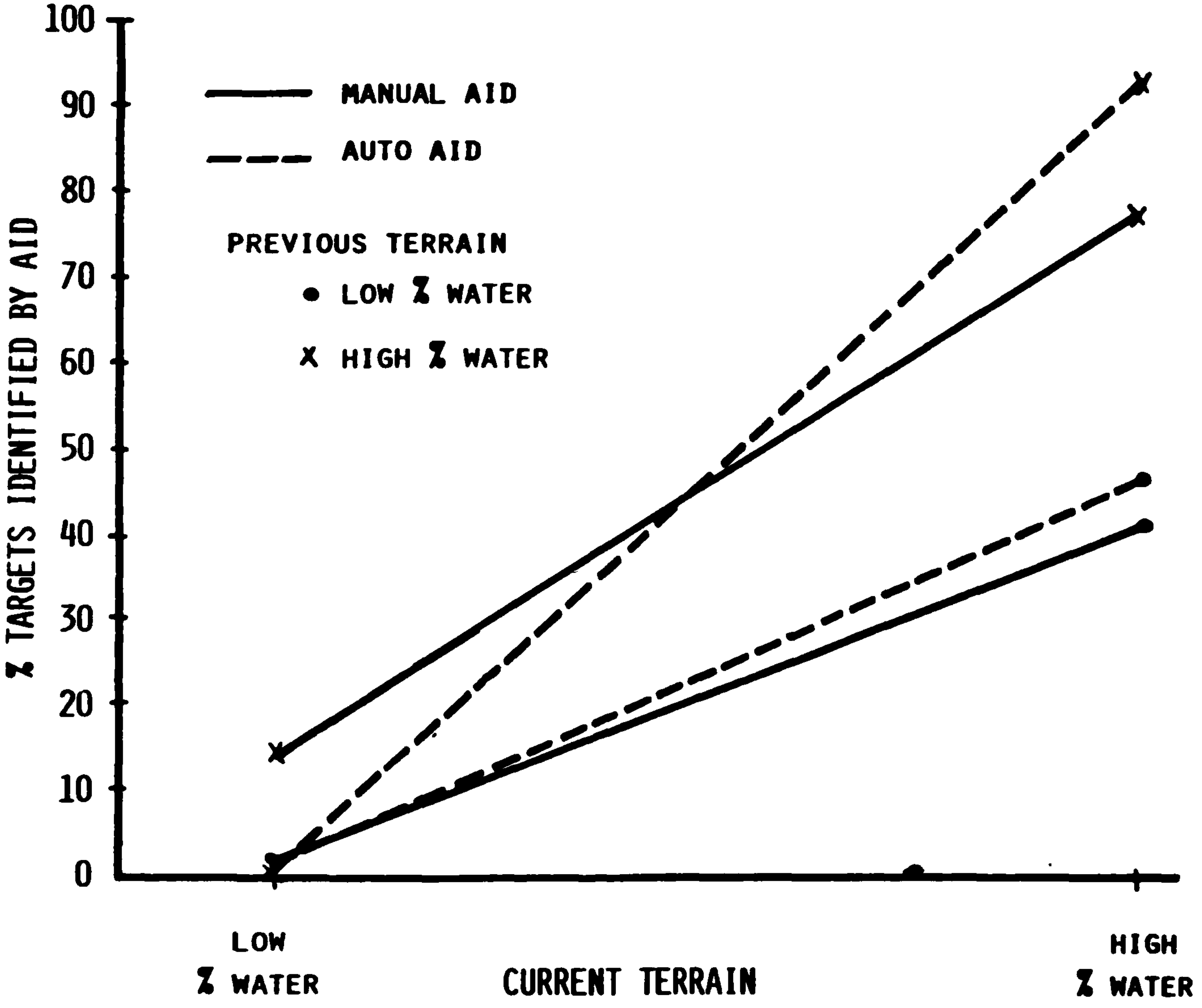


Figure 9. Targets identified by each type of aid.

both the previous and current terrain segments was high, the automatic aid achieved more hits than did the manual aid (93.89% vs. 77.18%).

Briefly summarizing patterns of manual vs. automatic aid activation, use of the automatic aid was more consistent and precise than use of the manual aid. The automatic aid turned itself on only over open water, and never identified targets in channels. In contrast, the manual aid occasionally identified targets in channels, and was sometimes not turned on over open water.

Effects of aiding upon unaided human performance. Recall that the results of the first experiment suggested that there might be beneficial effects of having an aid perform difficult portions of a task upon subsequent unaided performance of that task. In order to investigate this issue, spotting performance over terrain segments in which either the manual or automatic aid was available but not turned on was compared to performance over the same terrain segments when the aid was not available. First, it was noted that subjects identified more targets when the manual aid was available but turned off than in comparable conditions with no aid available (92.2% vs. 87.3%). This was consistent with the results of the first experiment. Interestingly, this was not true in the case of automatic aiding. When the automatic aid was turned off, subjects identified approximately the same number of targets as when no aid was available (88.5%). Differences were also noted in the latency of hits, which was lowest with manual aiding (21.67), higher with automatic aiding (23.02), and highest with no aid available (26.74).

Responses to Questionnaire Items

Responses to the questionnaire were compared to each other and to performance measures (i.e., rms error, hits, and false alarms) via paired t-tests and correlations. Discussion of significant correlations is confined to a subset of the large number which were obtained. It is felt that omission of the others from discussion is justified on the grounds that inclusion would not add to an understanding of the results of this experiment. For example, significant positive correlations between overall rms error, rms error over channels, and rms error over open water are excluded from discussion, as are negative correlations between estimated hits and false alarms.

Judgment of performance. Paired t-tests revealed the following discrepancies between estimated and actual task performance. First, subjective estimates of rms error over channels were lower than actual error (26.09 vs. 33.19). Second, estimates of hits over open water were higher than actual hits achieved (69.30% vs. 57.24%). Finally, overall false alarms were overestimated (0.63 vs. 0.20 actually achieved), as were false alarms over open water (0.94 estimated vs. 0.28 actually achieved).

Criteria for acceptable performance in self. Comparisons of "acceptable" performance ratings and subjective estimates of performance indicated that there were no significant differences. In other words, "acceptable" performance was roughly equivalent to how good subjects thought they actually were (which, as noted above, was sometimes better than they actually performed). Mean acceptable rms error was 30.12 (which would be achieved by reversing the direction of the tracking indicator just before it reached the yellow-red border). Achievement of 67.8% hits overall was indicated as acceptable, with 0.63 false alarms per terrain segment.

<u>Criteria for acceptable performance in assistants</u>. Generally, subjects indicated performance criteria for others which were more strict than those indicated for themselves. There were, however, no differences in performance requirements for human vs. computer assistants. In order for subjects to ask another person or computer for help, the assistant would have to achieve at least 86.25% hits with only 0.55 false alarms per terrain segment. Indicated criteria for acceptable percent hits on the part of assistants were also higher than subjects' estimates of their own overall performance (a mean of 73.40%).

<u>Attitudes and preferences about assistance</u>. Of the ten subjects, three preferred the automatic aid to the manual aid. The following reasons for this preference were given. One subject stated that he preferred the automatic aid because it freed him from the task of deciding when to turn the aid on and allowed him to concentrate on identifying boats. Another subject felt that

the automatic aid was more accurate with regard to hits and false alarms. When asked what he disliked about the other (i.e., manual) approach to aiding, one subject complained that there was too much to do and too little spare time when the manual aid was available.

Of the seven subjects preferring the manual approach to aiding, six preferred the manual aid because they wanted more control of the task. Two subjects indicated they had fewer false alarms with the manual aid. The following criticisms of automatic aiding were offered. Three subjects disliked the automatic aid because they did not know when the aid would transfer control of the spotting task, and transfers were sometimes disorienting. Two subjects disliked the lack of control and felt "secondary" to the computer. One subject disagreed with the computer's decisions and felt the aid was sometimes on too long, and one subject simply stated that the automatic aid made too many false alarms.

The difference between these two groups of subjects in the amount of time the manual aid was actually used (as indicated by the overall percent hits achieved by the manual aid) was not statistically significant due to variability across subjects (44.97% for those preferring the automatic aid, vs. 35.27% preferring the manual aid). However, when correlations between aid preference and other questionnaire items were computed, the following relationships were noted. (Aid preference was encoded as 1 for manual aid and 2 for automatic aid.) A negative relationship was observed between preference for the automatic aid and the percent of time automated decisions should agree with one's own in order for automatic decision making to be acceptable ($r = -.667$). There were also positive relationships between preference for the automatic aid and 1) indications of acceptable false alarms ($r = .802$), and 2) actual false alarms achieved over channels ($r = .742$).

Several other relationships were noted involving the percent of time automated decisions were required to agree with one's own in order to be acceptable. Negative correlations were observed with 1) acceptable false alarms ($r = -.844$) and 2) actual false alarms achieved over channels ($r = -.733$). Positive correlations were noted with 1) subjective estimates of percent hits achieved over water ($r = .701$), 2) percent hits required of a human assistant ($r = .638$), and 3) percent hits required of a computer assistant ($r = .867$).

Two relationships involving actual use of the manual aid were observed. As noted earlier, there was no difference in use of the aid by subjects preferring the manual vs. automatic aid. However, a negative relationship was noted between actual use of the manual aid and the degree to which an acquaintance was preferred over a stranger as an assistant ($r = -.633$). The relationship between use of the manual aid and actual hits achieved when no aid was available was also negative (r ranged from -.695 to -.727, dependent upon terrain composition). Thus, persons who used the aid less preferred acquaintances over strangers as assistants, and identified more targets themselves in unaided sessions.

Most of the relationships discovered in this analysis were intuitively reasonable. However, one relationship was observed which is counter-intuitive and puzzling. Negative correlations ranging from -.576 to -.659 were found between actual hits achieved and subjects' ratings of how "demanding" they felt they were. The way in which this relationship should be interpreted is not clear, and it is merely presented here as an intriguing result.

Summary of Results

- Consistent with the first experiment, there was an interaction of previous and current terrain types in their effects upon spotting performance (hits). Spotting was worse over bays than over channels, and carry-over effects of previous terrain type were observed.

- Minimal effects of spotting task difficulty on tracking performance were observed.

- Overall system performance improved when a spotting aid was available, with the greatest improvements when spotting over bays (i.e., when the aid was on). The availability of the aid reversed the impact of terrain composition on rms error (i.e., error was greater over channels), and overall hits were greater when an aid was available. There were also more false alarms, which could be attributed to the aid.

- Comparing performance with the two types of aid, spotting performance was better over channels than over bays when the manual aid was available; with the automatic aid available, spotting performance was better over bays.

- With the manual aid available, more false alarms occurred over channels than with either no aid or automatic aid available. There was no difference in false alarms over open water occurring with the two aided conditions, both of which were greater than false alarms with no aid available.

- The automatic aid consistently turned itself on only over open water; in contrast, use of the manual aid was less consistent, occasionally activated over channels and deactivated over bays.

- When subjects were in charge of the task allocation decision, their spotting performance when the aid was turned off was better than in comparable conditions in which no aid was available. When the task allocation decision was automatic, there was no such improvement when the aid was turned off. However, in those conditions in which the aid should have been used, system performance was better with the automatic aid.

- When asked to estimate their task performance, subjects underestimated tracking error over channels, overestimated hits achieved over bays, and overestimated false alarms overall and over bays.

- Subjects' ratings of "acceptable" performance in themselves were approximately equal to their estimates of performance achieved. Ratings of acceptable performance in an assistant (human or computer) were more demanding, with more hits and fewer false alarms required in order to consider using the assistant's help.

- Seven of ten subjects preferred the manual approach to aiding, citing reasons of desire to be in control, lack of understanding of when the automatic aid transferred control, disagreement with the automatic aid's decisions, and feeling "secondary to the computer" when it made task allocation decisions.

- Three subjects preferred the automatic aid, reporting that it freed them from having to make the allocation decision, allowed for more spare time, and was more accurate with respect to hits and false alarms.

- Preference for the automatic aid was negatively correlated with ratings of the degree to which the computer's task allocation decisions should agree with one's own decisions, and positively related to the extent to which the computer's performance resembled one's own performance and performance criteria (i.e., with regard to false alarms).

- Required level of agreement between the computer's decisions and one's own decisions was positively related to the extent to which the computer's performance resembled one's own performance and performance criteria, and negatively related to estimates of the quality of one's own performance and criteria for performance in an assistant.

- Use of the manual aid was positively correlated with preference for an acquaintance as an assistant, and negatively related to the quality of unaided spotting performance.
- Subjects who rated themselves as more demanding achieved fewer hits than less demanding subjects.

DISCUSSION

A number of comments may be made about the results of the research reported here. First, the parameters of the tracking and spotting tasks affected performance in anticipated ways. In short, the difficult manipulations were successful. One effect which was not expected was the carry-over effect of terrain type on spotting performance, with current spotting performance affected by the amount of water in the previous terrain segment. This effect was observed in both experiments, and appears to be quite robust in this environment.

An intuitive explanation for the effect is that subjects spotting over a broad area of water had to focus more on the current terrain and were unable to preview the upcoming terrain. As a result, their performance in the next terrain segment was not as good as it could have been if they had been able to look ahead. Intuitive though it may be, the notion that increased task difficulty can shorten one's planning horizon and consequently affect future performance is consistent with some laboratory studies and anecdotal evidence from a variety of domains (Johannsen & Rouse, 1979, 1983). The lesson to be learned here is that performance on a task may depend not only on current conditions but also on what one has just finished doing.

Contrary to expectations, the relationship between performance of the two tasks was not very strong. It seems plausible that this failure to note clear tradeoffs in performance can be partially attributed to the nature of the tracking task. Performance of the tracking task was rather simple, requiring only bang-bang control via the space bar on the terminal keyboard. Performance was also necessary, since targets could not be identified if the controlled element was out-of-bounds. Stronger effects of tracking difficulty on spotting performance and vice versa might have been noted if continuous control had been required or if the option of "shedding" the tracking task had been available.

An additional explanation for the weak relationship noted may be found in the criterion for acceptable performance apparently adopted by subjects. Most subjects seemed to accept considerable error, merely keeping the position indicator out of the red region. This choice of criterion, which was also indicated by responses to the questionnaire, was reasonable from the subjects' point of view, since there was no penalty associated with tracking error other than an inability to identify targets if the indicator was in the red region. The result, however, was a compressed range of rms error scores, which may have obscured any differences due to spotting task parameters. There is also the strong possibility that there were in fact few differences, because performance of the tracking task was not sufficiently demanding to affect spotting performance.

The hypothesis that adaptive aiding can lead to improvements in system performance is supported by the positive effects of having an aid available to perform the spotting task. For example, performance on both tasks improved when the spotting aid was available. There was also less performance tradeoff between tasks (i.e., performance of one task was less affected by the difficulty of the other task).

This research also revealed that subtle effects may occur as well. For example, it was noted unexpectedly that having an aid available to perform the difficult portions of a task may also enhance unaided performance. It was even more surprising to discover that attainment of this benefit occurred in this research only when subjects turned the aid on and off themselves. The reasons for this are not clear, and await exploration.

When subjects were in charge of spotting task allocation, activation of the aid was less consistent than when the decision was automated. Lack of consistency alone is not necessarily indicative of poor decision making on the part of the human, however, but rather could reflect variability in need for assistance. Recall that subjects' average use of the aid was appropriate to their average need. However, the fact that spotting performance over water was better when the aid made the task allocation decisions suggests that subjects' decisions over water were not as appropriate as the aid's decisions.

Thus, neither approach to aiding was clearly superior to the other, and each had unique benefits to offer. Rather than answering questions about how adaptive aiding should be implemented, these results underscore subtleties and complicate the issue. When the human is able to summon help as desired, his/her own performance may improve, but the full benefits of assistance may not be realized because the assistant is not summoned as frequently as it should be. The behavior of one subject reminds us that inappropriate over-reliance upon the aid is possible also.

One approach to resolving this dilemma might be to identify ways to enhance the quality of decisions made by both the human and computer. For the computer, better models to serve as the basis for decisions are needed. The results presented here offer a few implications for modeling efforts. For example, it may be necessary to test the validity of models for predicting performance in aided contexts by examining unaided performance within the aided context. Additionally, if performance may be expected to change over time, there should be some mechanism for adjusting the model to accommodate these changes.

There are several intuitively reasonable candidates for factors influencing the quality of the human's decisions. These are elaborated in the conceptual framework (Morris, Rouse, & Frey, 1985), and include factors such as motivation, attitudes toward the aid, and need to be in control. One reason that subjects in this research did not use the aid as often as they should have may have been that they did not think they needed help. Recall that estimates of spotting performance over water were higher than actual performance achieved.

Referring again to the conceptual framework, information available to the human is viewed as an important contributor to judgments of one's own performance and to the quality of task allocation decisions. In light of these results and the central role information is given in the conceptual framework, the nature of information required by the human to make good decisions will be the next focus of this research. Incidentally, investigation of information requirements may provide a clue as to why unaided performance did not improve when the automatic aid was available. Some subjects reportedly did not like the automatic aid because they did not understand when it would turn itself on or off and found it disconcerting.

Two statements may be made about attitudinal factors, based on responses to the questionnaire. These statements are merely suggested by the data, and it is anticipated that future results will allow refinement and specification of limiting conditions. They are presented here as "straw men" to be tested. First, if task accuracy is important to a person, he/she will not want to surrender control of that task to an aid unless the aid is perceived as substantially better than himself/herself. Second, the more similar an aid's performance is to a person's performance, the less it matters to that person whether or not the aid's task allocation decisions agree with his/her own decisions. In other words, the less impact differences in decisions will have on overall performance, the less those differences matter. The second statement was based primarily on relationships observed involving false alarms; thus, the following alternative interpretation of these relationships is offered. An aid's decisions may disagree with the human, as long as it is unlikely the aid will do anything wrong in performing its task (e.g., make a false alarm).

In spite of all best efforts, it is virtually inevitable that a situation will arise in which computer and human disagree. This prompts a very important question: What should be done if the human and model disagree? Under what conditions should the human prevail, and when should the human be "saved from himself"? The answer to this question is not at all straightforward. Even in this simple task environment, it was observed that humans were sometimes right and sometimes wrong in disagreeing with the model used.

No attempt is made to answer this question here. Rather, it is pointed out that the answer depends on a number of practical, ethical, and philosophical issues, such as the frequency with which the human and model may be expected to disagree, consequences of error on the part of the human and model, and one's position on question of which partner should ultimately be "in charge" of the system. Dependent on conditions, a variety of approaches to aiding may have to be employed. Two approaches were used in this research: either the human requested the aid's assistance when it was desired, or the computer made the task allocation decisions without offering the human any recourse. Alternatively, the aid could: 1) suggest itself but do nothing unless the human indicated acceptance of the suggestion, 2) or perform a task unless overridden by the human.

FUTURE DIRECTIONS

As noted, the next focus of this research will be investigation of information required by the human about an aid in order to make effective decisions about the use of that aid. An "armchair" analysis has been conducted, and is presented in a working paper (Morris & Rouse, 1985). Pursuit of this topic will involve expansion of the experimental task environment to include a wider variety of tasks, and elaboration of the aid's task performance. The effects of various types of information on the quality of the human's decisions will be investigated in both familiar and novel situations.

APPENDIX:

QUESTIONNAIRE USED IN EXPERIMENT TWO

END-OF-EXPERIMENT QUESTIONNAIRE

Your answers to the following questions about the experiment and your preferences about aiding will be greatly appreciated. Questions about your performance refer to your performance <u>without</u> the aid available.

1. On the average, how many of the target boats present do you think you identified? (Indicate by placing an X on the scale.)

   0% : 20% : 40% : 60% : 80% : 100% :

2. How many of the targets did you identify in the areas which were mostly land?

   0% : 20% : 40% : 60% : 80% : 100% :

3. How many of the targets did you identify in the areas which were mostly water?

   0% : 20% : 40% : 60% : 80% : 100% :

4. On the average, how many false alarms did you make in one pass over the terrain (relative to hits)?

   none : : half as many as hits : : as many as hits :

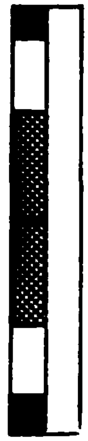
5. How many false alarms did you usually have over land?

   none : : half as many as hits : : as many as hits :

6. How many false alarms did you usually have over water?

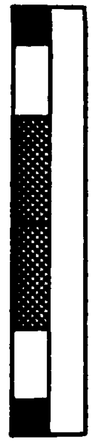
   none : : half as many as hits : : as many as hits :

7. On the average, within what range do you think you maintained the tracking task? (Indicate upper and lower limits on figure 7.)



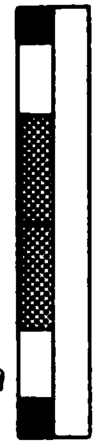
7

8. Within what range did you keep the tracking task when you were spotting over land? (Indicate on figure 8.)



8

9. Within what range did you keep the tracking task when you were spotting over water? (Indicate on figure 9.)



9

Obviously, since the tasks used in this experiment were "artificial", your task performance in this experiment had no bearing on anything outside of the laboratory. However, we hope that information gained from this experiment will be helpful in future real-world situations. Therefore, please try to answer the following questions as if obtaining an accurate estimate of water traffic was actually important to you.

10. Independent of the amount of water in the window, what is the worst performance you would consider to be acceptable? (That is, if you performed at least as well as this, you would be satisfied with your performance.)

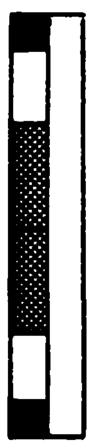
HITS:

0% : 20% : 40% : 60% : 80% : 100% :

FALSE ALARMS:

none : | : | half as many as hits : | : | as many as hits :

11. Independent of the amount of water in the window, what is the maximum range on the tracking task which you would consider acceptable? (That is, if you performed as well as this, you'd be satisfied with your performance.) Indicate on figure 11.



11

12. Suppose another <u>person</u> was available to help you by performing the spotting task. How good would that person have to be for you to consider asking him/her for help over land?

HITS:

0% 20% 40% 60% 80% 100%

: : : : : :

FALSE ALARMS:

none — half as many as hits — as many as hits

: : : : :

☐ I would not consider asking for help over land.

13. How good would that person have to be for you to consider asking him/her for help over water?

HITS:

0% 20% 40% 60% 80% 100%

: : : : : :

FALSE ALARMS:

none — half as many as hits — as many as hits

: : : : :

☐ I would not consider asking for help over water.

14. Independent of your assistant's skill at the spotting task, would it make a difference in your willingness to accept help if you knew your helper, as opposed to working with a stranger?

strongly prefer stranger — no preference — strongly prefer acquaintance

: : : : :

Why? ____________________

____

____

15. Now, suppose a computer was available to do the spotting task rather than a person (as it was in some cases during the experiment). How well would the computer have to be able to do the spotting task for you to consider asking it for help over land?

HITS:

0%	20%	40%	60%	80%	100%
:	:	:	:	:	:

FALSE ALARMS:

none		half as many as hits		as many as hits
:	:	:	:	:

☐ I would not consider asking for help over land.

16. How well would the computer have to perform for you to consider asking it for help over water?

HITS:

0%	20%	40%	60%	80%	100%
:	:	:	:	:	:

FALSE ALARMS:

none		half as many as hits		as many as hits
:	:	:	:	:

☐ I would not consider asking for help over water.

17. If you had your choice between a computer or a person, with equal performance characteristics (i.e., scored the same number of hits and false alarms), which would you prefer to help you?

strongly prefer computer		no preference		strongly prefer person
:	:	:	:	:

Why?____

____

____

18. How much better would the <u>other</u> helper have to be for you to prefer it over the one you chose?

1% better		50% better		100% better or more
:	:	:	:	:

☐ I would never choose the other helper.

If you would never choose the other helper, why not?

____

____

____

In this experiment, different approaches were used when the computer aid stepped in.

1) Sometimes the computer made all of the decisions, without giving you the opportunity to override it.

2) At other times, you were the decision maker, and the computer never did anything unless you requested it.

The following questions refer to these different approaches to aiding.

19. Which of the approaches to aiding did you like better? Why?

____

____

____

20. What did you dislike about the other approach to aiding?

____

____

____

21. Suppose the first approach to aiding was to be used in a real system (that is, the computer was to make all decisions as to who should do the spotting). How closely would the computer's decisions have to agree with what you would do for you to feel comfortable about the computer being the decision maker? (Indicate percent agreement.)

0% 20% 40% 60% 80% 100%

☐ I would never feel comfortable with the computer making the decisions.

22. If you only received the output of the computer's performance (that is, hits and false alarms) and could not watch as the computer performed the spotting task, would that change your answer to question 21? If so, how and why?

____

____

____

23. Suppose the computer was the decision maker, but you could override its decisions if you wished. For example, when the computer informed you that it was about to take over or give the spotting task to you, you could override it by pressing a button on the mouse, and control of the spotting task would not be transferred. How closely would the computer's decisions have to agree with yours for you to feel comfortable about the computer being the decision maker? (Indicate percent agreement.)

0% 20% 40% 60% 80% 100%

☐ I would never feel comfortable with the computer making the decisions.

24. Forget about the tasks performed in this experiment for the moment, and think more broadly about the kinds of tasks you are usually responsible for (such as school projects or things you do at your job). Assuming it was "OK" to delegate work, and someone was available who could do the work to your satisfaction, how likely would you be to have someone else do some of your work for you?

extremely unlikely : ______ : ______ : ______ : ______ : extremely likely

If you would not delegate work to someone else, why not?

____

____

____

25. In general, how easy is it to find people who perform work to your satisfaction? (In other words, how demanding are you?)

very easygoing : ______ : ______ : ______ : ______ : very demanding

REFERENCES

Jex, H. R., McDonnell, J. P., and Phatak, A. V. (1966). A "critical" tracking task for manual control research. IEEE Transactions on Human Factors in Electronics, HFE-7, 138-144.

Johannsen, G., & Rouse, W. B. (1979). Mathematical concepts for modeling human behavior in complex man-machine systems. Human Factors, 21, 733-747.

Johannsen, G., & Rouse, W. B. (1983). Studies of planning behavior of aircraft pilots in normal, abnormal, and emergency situations. IEEE Transactions on Systems, Man, and Cybernetics, SMC-13, 267-278.

Morris, N. M., & Rouse, W. B. (1985). Information requirements for effective use of adaptive aiding: Who's in charge, and what do they need to know? (working paper). Search Technology, Norcross, GA.

Morris, N. M., Rouse, W. B., & Frey, P. R. (1985). Adaptive aiding for symbiotic human-computer control: Conceptual model and experimental approach (Tech. Report AFAMRL-TR-84-072). Search Technology, Norcross, GA.

Rouse, W. B., & Rouse, S. H. (1983). A framework for research on adaptive decision aids (Tech. Report AFAMRL-TR-83-082). Search Technology, Norcross, GA.

* U.S. GOVERNMENT PRINTING OFFICE 1986 - 646-006/200247

END

DTIC

5-86